DOCUMENT RESUME

ED 244 996 TM 840 320

AUTHOR Drudge, Owen W.; And Others
TITLE A Comparison of the WISC-R, McCarthy Scales, Woodcock-Johnson, and Academic Achievement: Concurrent and Predictive Validity.
PUB DATE [81]
NOTE 20p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Achievement Tests; *Cognitive Tests; Correlation; Grade 1; *Intelligence Tests; Predictive Validity; Primary Education; Student Evaluation; *Test Validity
IDENTIFIERS *McCarthy Scales of Childrens Abilities; Wechsler Intelligence Scale for Children (Revised); *Woodcock Johnson Tests of Cognitive Ability

ABSTRACT

Thirty normal first-grade students with a mean age of 7.0 years were administered the Wechsler Intelligence Scale for Children-Revised, the McCarthy Scales of Children's Abilities, and the Woodcock-Johnson Tests of Cognitive Ability. Pearson intercorrelations among the summary indices of these cognitive measures were uniformly high, ranging from .78 to .94. Mean differences were not found between the three full scale indices. However, the short form index of the Woodcock-Johnson was found to be a consistent, significant underestimate of the full scales. Academic achievement was measured by the Wide Range Achievement Test and teacher ratings. Correlations between the cognitive and achievement measures were significant in all cases. Although striking differences were not apparent, the Woodcock-Johnson indices proved to correlate slightly better with achievement. Partial correlations between the Wide Range Achievement Test and teacher ratings were likewise all significant and were ordered in a manner that would suggest that the two sets of scales measured similar domains. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED244996

A Comparison of the WISC-R, McCarthy Scales, Woodcock-Johnson, and Academic Achievement: Concurrent and Predictive Validity[1]

Owen W. Drudge, Thomas P. Reilly, James C. Rosen,

Mariellen Fischer, and Dorene Loew

University of Vermont

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

O. W. Drudge

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running head: Cognitive Ability and Academic Achievement

TM 840 320

Abstract

Thirty normal first-grade students with a mean age of 7.0 years were administered the Wechsler Intelligence Scale for Children - Revised, the McCarthy Scales of Children's Abilities, and the Woodcock-Johnson Tests of Cognitive Ability. Pearson intercorrelations among the summary indices of these cognitive measures were uniformly high, ranging from .78 to .94. Mean differences were not found between the three full scale indices. However, the short form index of the Woodcock-Johnson was found to be a consistent, significant underestimate of the full scales. Academic achievement was measured by the Wide Range Achievement Test and teacher ratings. Correlations between the cognitive and achievement measures were significant in all cases. Although striking differences were not apparent, the Woodcock-Johnson indices proved to correlate slightly better with achievement. Partial correlations between the Wide Range Achievement Test and teacher ratings were likewise all significant and were ordered in a manner that would suggest that the two sets of scales measured similar domains.

A Comparison of the WISC-R, McCarthy Scales, Woodcock-Johnson, and Academic Achievement: Concurrent and Predictive Validity

Adequate diagnostic and descriptive psychological assessment of school-age children is limited largely by the integrity of the psychometric test, or tests, employed in the evaluation process. For many years the Wechsler Intelligence Scale for Children (WISC) and its revision, the Wechsler Intelligence Scale for Children - Revised (WISC-R), have been considered acceptable instruments for measuring intellectual functioning in primary and secondary school populations (Wechsler, 1949; 1974). Although the WISC-R offers a reasonable estimate of global intelligence, it alone frequently fails to furnish sufficient detail regarding the examinee's particular strengths and weaknesses relevant to psychoeducational assessment. The McCarthy Scales of Children's Abilities (MSCA) (McCarthy, 1972) and the Woodcock-Johnson Tests of Cognitive Ability (WJTCA) (Woodcock & Johnson, 1977) are two newer measures which have been found to be useful adjuncts and/or alternatives to the WISC-R. The MSCA was constructed with a developmental emphasis and has standardized norms for ages 2½ to 8½ years, making it appropriate for preschool-age and early primary school-age children. It consists of six scales which measure motor ability and a range of cognitive skills including verbal, quantitative, perceptual-performance, and memory abilities. The WJTCA, an instrument designed specifically for psychoeducational purposes, consists of four scholastic aptitude scales and four cognitive ability scales. In addition to their various special scales, the MSCA and WJTCA each offer a summary

score or index of overall cognitive functioning; these are the General Cognitive Index and the Full Scale Broad Cognitive Ability, respectively. Both scales are similar to the Full Scale IQ of the WISC-R in that they are estimates of general cognitive, and perhaps intellectual, ability and are derived from most or all of the subtest scores of the instruments.

Many studies evaluating the concurrent and predictive validity of the WISC-R have found the test to correlate reasonably well with other measures of cognitive ability, measures of academic achievement, and related performance variables (e.g., Brooks, 1977; Covin & Lubimiv, 1976; Hartlage & Steele, 1977; Schwarting & Schwarting, 1977). Research comparing the WISC-R to the MSCA (Gon & Youngquist, 1979; Ivimey & Taylor, 1980; Naglieri, 1980a) and to the WJTCA (Reeve, Hall, & Zakreski, 1979; Woodcock, 1978; Ysseldyke, Shinn, & Epps, 1981) consistently has found both tests to have moderate to high concurrent validity with the WISC-R. Studies contrasting the relationship of these tests with academic achievement are decidedly few in number and have examined only the correlations between these instruments and performance on an achievement test (Ivimey & Taylor, 1980; Naglieri, 1980b) or tests (Woodcock, 1978), ignoring other possible and appropriate criteria such as school grades or teacher ratings of academic achievement. Thus far there have been no validation studies examining the relationship between the MSCA and the WJTCA.

The purpose of the present study was to generate more extensive validity information and replicate, or partially replicate, some of the previous studies by: (a) comparing the full scale or summary cognitive indices of the WISC-R, MSCA, and WJTCA directly to one another, and (b) comparing the relationship of these measures with performance on a

standardized test of academic achievement and with teacher ratings of student achievement based on classroom performance.

Method

Subjects. First-grade students were recruited from four regular classrooms in two local elementary schools. The parents of all first-graders in those classes received letters describing the study and soliciting consent for their child's participation. Consent was granted for 33 (42%) of the 78 available students. Because scheduling arrangements could not be made for three children, only 30 (38%) subjects were included in the study. Those that participated were all Caucasian, came from a full range of socioeconomic backgrounds, and were divided equally by sex. Their ages ranged from 6.5 years to 8.1 years at the date of testing with a mean of 7.0 years (SD = 0.4).

Procedure. At the end of their first-grade school year, each of the subjects was tested with the WISC-R, MSCA, WJTCA, and the Wide Range Achievement Test (WRAT) (Jastak & Jastak, 1978). The order of administration for the three cognitive ability measures was counterbalanced. Tests were administered by five examiners, consisting of one Ph.D. psychologist and four pre-doctoral psychology graduate students. The testing of each subject occurred over a 1-3 day period and, in almost all cases, all four tests were administered to that subject by the same examiner. As a separate and independent measure of academic achievement, teacher ratings were obtained for each student on a five point scale (1 = Poor; 2 = Below average; 3 = Average; 4 = Above average; 5 = Superior) for the same three academic

domains assessed by the WRAT: Reading, Spelling, and Arithmetic. These ratings were the only measure of classroom performance available because both schools participating in the study employed the equivalent of a pass/fail system for their younger primary school students. Test forms for the various measures were not scored until after all data collection was completed to help minimize possible inter-test contamination through examiner expectation or bias.

Analyses. Pearson intercorrelations were calculated to determine the degree of similarity or common variance between the Full Scale IQ (FSIQ) of the WISC-R, the General Cognitive Index (GCI) of the MSCA, and the Full Scale and the Brief Scale Broad Cognitive Ability (FSBC and BSBC) of the WJTCA. The FSBC is derived from all twelve of the Woodcock-Johnson cognitive subtests, whereas only two of the subtests are included in the abbreviated scale, the BSBC. The means of these four cognitive indices (i.e., FSIQ, GCI, FSBC, and BSBC) were compared by a repeated measures ANOVA and Student-Newman-Keuls a posteriori contrasts between the means of all index pairs. The relationship of the indices with achievement was evaluated by Pearson correlations for the WRAT subscales and by partial correlations adjusted for age for the teacher achievement ratings. The age correction for the ratings was necessary because the teachers were asked to evaluate each student in reference to all of his or her first-grade classmates and not just to same-age peers. Partial correlations were also used to examine the relationship between the WRAT subscale scores and teacher ratings. The significance of the differences between correlation pairs (e.g., FSIQ with WRAT Reading v. GCI with WRAT Reading) was determined through Hotelling's (1931) t test adaptation for dependent

correlational pairs.

Results

The intercorrelations, means, and standard deviations for the four cognitive indices are presented in Table 1. All of the Pearson correlations between the indices were significant and ranged from .78 to .94, suggesting appreciable common or shared variance. The repeated measures ANOVA yielded a significant F ratio for between measures ($F = 5.57$; $df = 3,57$; $p < .002$). The Student-Newman-Keuls a posteriori contrasts found no differences between the FSIQ, GCI, and FSBC. However, the Brief Scale of the Woodcock-Johnson, the BSBC, was found to be a consistent and significant ($p < .05$) underestimate of the three full scales.

Insert Table 1 about here

The correlations between cognitive ability and the measures of achievement (WRAT and teacher ratings) were likewise all significant, ranging from .45 to .74 and .53 to .77, respectively (see Table 2). In nearly all cases, the FSBS and BSBC of the WJTCA correlated better than the FSIQ and GCI with the various achievement subscales. Although reasonably consistent, the magnitude of these differences was often slight. Because of the large number of t tests necessary to contrast all correlational pairs of each index with the various achievement subscales, 48 in all, a conservative alpha-level of .01 was employed for the analyses to minimize the possiblity of spurious findings. Despite the apparent trend of FSBC and/or BSBC being greater than than the FSIQ and GCI, none of these differences observed between the

correlation coefficients were found to be significant at the alpha = .01 level.

Insert Table 2 about here

Partial correlations comparing the WRAT and the teacher ratings, presented in Table 3, were also significant in every case and ranged between .54 and .84. Correlations derived from all subscale combinations excluding the WRAT Arithmetic subscale were uniformly high, while those that included it tended to be slightly smaller in magnitude.

Insert Table 3 about here

Discussion

A number of relevant findings are apparent within these data. First, the similarity of the means of the WISC-R FSIQ, McCarthy GCI, and Woodcock-Johnson FSBC and the magnitude of their intercorrelations suggest that, at least for normal or non-clinical groups, these indices are extremely comparable. Previous studies examining similar populations have reported equivalent findings for the relationship between the FSIQ and GCI (Davis & Walker, 1977; Harrison & Wiebe, 1977). Neither of these studies found differences between the means for the two scales and reported correlation coefficients of .75 and .74, respectively, which are less, but not greatly dissimilar from the present study's .86. In comparing the WISC-R FSIQ to the FSBC as part of the WJTCA standardization and development, Woodcock (1978) found

the two indices to correlate .79 for third-grade and fifth-grade groups. Once again, this is slightly smaller than the correlation of .89 observed for these scales in the present study. As mentioned earlier, no prior studies have examined the relationship of the GCI and FSBC. Considering that these scales are derived from very different tests, their .83 correlation and the equivalence of their means suggest they have similar psychometric properties. Also worth noting is the discrepancy between the three full scale indices (the FSIQ, GCI, and FSBC) and the Woodcock-Johnson BSBC. This brief index correlates well with all three, with coefficients ranging from .78 to .94, but for reasons yet unclear, it is a significant underestimate of each, including the FSBC with which it shares common subtests. This discrepancy apparently has not been observed previously and warrants further investigation.

Perhaps of most interest is the relationship between the cognitive measures and achievement. In this study, all correlations between the four cognitive indices and the objective and subjective measures of academic achievement were statistically significant and of at least moderate magnitude. Comparisons of the various Pearson and partial correlations failed to clearly identify one cognitive measure as being superior to the others. The FSBC and BSBC of the Woodcock-Johnson most often produced the largest correlations with the WRAT and teacher ratings, but the size of these differences was generally modest and accounted for only little additional variance. Woodcock (1978) reported a similar correlational relationship for the two WJTCA indices and the WISC-R Full Scale IQ with academic achievement. He found that the FSBC and BSBC were almost always minimally better

predictors of performance on a number of achievement tests, including the WRAT; the magnitude of those differences were very much like those presented here in Table 2. Taken together, Woodcock's data and the findings of this present study offer some support for the contention that, relative to the WISC-R, the WITCA is more a test of scholastic aptitude than of intellectual capacity and, therefore, should be more closely related to achievement (see Ysseldyke et al., 1981).

The decision to obtain teacher ratings proved to be both interesting and valuable. Although only global estimates of classroom performance, these ratings correlated well with their corresponding WRAT subtests, suggesting that the two sets of scales do seem to measure similar domains. The ratings may have in fact been a better measure of arithmetical achievement as the teacher scores tended to correlate more highly with cognitive ability than did the WRAT. Compared to the WRAT Reading and Spelling subtests, the range of scores obtained by subjects on the Arithmetic subtest was clearly constricted, as can be inferred by their standard deviations presented in Table 2. It is quite probable that this lack of variability was responsible, at least in part, for the rather deflated correlation coefficients obtained. For his third-grade population, Woodcock (1978) also found markedly lower correlations for the WRAT Arithmetic subtest when comparing it to other tests of mathematical achievement. However, this was not the case for his older groups for the Arithmetic subtest or for the WRAT Reading and Spelling subtests with the third-graders, where correlations comparable to other achievement tests were produced. It appears, therefore, that the problem that exists is largely specific to the use of the Arithmetic subtest with younger individuals, such as those in the present study.

The findings reported here would indicate that the FSIQ of the WISC-R, the GCI of the McCarthy Scales, and the FSBC and BSBC of the Woodcock-Johnson are reasonably comparable to one another in their concurrent properties. However, these data should be interpreted within the context of the population examined and generalized to other populations with a measure of caution. The students selected for this study were recruited from four general classrooms and are likely representative of a "normal" population, with two possible exceptions. First, due to a paucity of ethnic and minority groups residing in the greater geographical locale of this study, the sample derived was unsurprisingly completely Caucasian. Generalization of findings such as these to minority groups is yet controversial and should be done only with sufficient consideration and discretion (Cole, 1981; Olmedo, 1981). Secondly, for reasons uncertain, the group means for the cognitive indices obtained here were several scale points above the means established in the standardization norms for these tests (i.e., 50th percentile = 100). Whereas this could result from having a disproportionate number of higher functioning students selected from the various classrooms, the means of the teacher ratings argue against this and in fact suggest that the tested sample was essentially average relative to the rest of their peers. In any event, these elevations in index scores should not significantly affect the overall correlational relationship between these measures, nor should they appreciably detract from the utility of the findings of this study in general.

The group used here was younger than those reported in other comparative validation studies involving these measures. This is important because not only does the present study offer new validation information

in the form of novel comparisons between tests, but it also extends the existing knowledge of these instruments to a population that has been somewhat under studied. Given the recent trend towards early assessment and identification of children requiring special school services, the need for the development of reliable and valid tests for use with the younger populations cannot be overemphasized. Specifically, there is a necessity for future research to address the diagnostic validity of these tests. Previous studies with learning disabled populations have found that the McCarthy GCI (Ivimey & Taylor, 1980; Naglieri, 1980a) and the FSBC of the Woodcock-Johnson (Reeve et al., 1979; Ysseldyke et al., 1981) both tend to underestimate the WISC-R FSIQ. Most guidelines established for educational diagnosis and program assignment, such as for the learning disabled, are based on fairly precise estimates of an individual's cognitive or intellectual ability. Therefore, if indeed the GCI and FSBC are eventually proved to be discrepant predictors of cognitive functioning, as suggested by these early studies, the role of these instruments in the assessment and diagnostic process necessarily will be diminished, but perhaps not dramatically so. Such findings would impact primarily on their viability as possible replacements for the traditional IQ tests in diagnostic evaluations. However, the status of these tests as being valuable adjuncts to the WISC-R, or as being acceptable measures for assessments where a specific diagnosis in not sought, seemingly should remain unscathed. Actually it would be quite marvelous if future research was to confirm the presence, and reliability, of score discrepancies between the FSIQ and these two measures and further, that these discrepancies were found to be specific to the learning disabled population. Patho-

gnomonic indicators are rare in psychological assessment and should be appreciated in whatever form they present themselves. In closing, there remains little real doubt that the McCarthy Scales of Children's Abilities and the Woodcock-Johnson Tests of Cognitive Ability are both valuable and appropriate tests for use with school-age children. How they best can be utilized, however, has yet to be determined.

References

Brooks, C. R. WISC, WISC-R, S-B L &M, WRAT: Relationships and trends among children ages six to ten referred for psychological evaluation. Psychology in the Schools, 1977, 14, 30-33.

Cole, N. S. Bias in testing. American Psychologist, 1981, 36, 1067-1077.

Covin, T. M., & Lubimiv, A. J. Concurrent validity of the WRAT. Perceptual and Motor Skills, 1976, 43, 573-574.

Davis, E. E., & Walker, C. McCarthy Scales and WISC-R. Perceptual and Motor Skills, 1977, 44, 966.

Goh, D. S., & Youngquist, J. A comparison of the McCarthy Scales of Children's Abilities and the WISC-R. Journal of Learning Disabilities, 1979, 12, 344-348.

Harrison, K. A., & Wiebe, M. J. Correlational study of McCarthy, WISC, and Stanford-Binet scales. Perceptual and Motor Skills, 1977, 44, 63-68.

Hartlage, L. C., & Steele, C. T. WISC and WISC-R correlates of academic achievement. Psychology in the Schools, 1977, 14, 15-18.

Hotelling, H. The generalization of Student's Ratio. Annals of Mathematical Statistics, 1931, 2, 360-378.

Ivimey, J. K., & Taylor, R. L. Differential performance of learning disabled and non-learning disabled children on the McCarthy Scales, WISC-R, and WRAT. Journal of Clinical Psychology, 1980, 36, 960-963.

Jastak, J. F., & Jastak, S. The Wide Range Achievement Test: Manual of instuctions. Los Angeles: Western Psychological Services, 1978.

McCarthy, D. Manual for the McCarthy Scales of Children's Abilities. New York: Psychological Corporation, 1972.

Naglieri, J. A. McCarthy and WISC-R correlations with WRAT achievement scores. Perceptual and Motor Skills, 1980, 51, 392-394. (a)

Naglieri, J. A. Comparison of McCarthy General Cognitive Index and WISC-R IQ for educable mentally retarded, learning disabled, and normal children. Psychological Reports, 1980, 47, 591-596. (b)

Olmedo, E. L. Testing linguistic minorities. American Psychologist, 1981, 36, 1078-1085.

Reeve, R. E., Hall, R. J., & Zakreski, R. S. The Woodcock-Johnson Tests of Cognitive Ability: Concurrent validity with the WISC-R. Learning Disabilities Quarterly, 1979, 2, 63-69.

Schwarting, F. G., & Schwarting, K. R. The relationship of the WISC-R and WRAT: A study based upon a selected population. Psychology in the Schools, 1977, 14, 431-433.

Wechsler, D. Wechsler Intelligence Scale for Children. New York: Psychological Corporation, 1949.

Wechsler, D. Wechsler Intelligence Scale for Children - Revised. New York: Psychological Corporation, 1974.

Woodcock, R. W. Development and standardization of the Woodcock-Johnson Psycho-Educational Battery. Boston: Teaching Resources, 1978.

Woodcock, R. W., & Johnson, M. B. Woodcock-Johnson Psycho-Educational Battery. Boston: Teaching Resources, 1977.

Ysseldyke, J., Shinn, M., & Epps, S. A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability. Psychology in the Schools, 1981, 18, 15-19.

Footnotes

[1] The authors wish to offer special thanks to the teachers, principals, and other personnel of the Burlington School District for their support, cooperation, and patience instrumental in the development and completion of this study.

Table 1

Intercorrelations, Means, and Standard Deviations for the Four Cognitive Indices

	FSIQ	GCI	FSBC	BSBC
BSBC	.85[a]	.78	.94	---
FSBC	.89	.83	---	---
GCI	.86	---	---	---
Mean[b]	108.2	109.6	110.5	104.7
S.D.	14.5	16.1	14.5	14.8

Note: n = 30. FSIQ = Full Scale IQ (Wechsler Intelligence Scale for Children - Revised); GCI = General Cognitive Index (McCarthy Scales of Children's Abilities); FSBC = Full Scale Broad Cognitive Ability, BSBC = Brief Scale Broad Cognitive Ability (Woodcock-Johnson Tests of Cognitive Abilities).

[a] $p < .001$ for all correlation coefficients.

[b] ANOVA: $p < .002$; Student-Newman-Keuls: FSIQ = GCI = FSBC > BSBC.

Table 2

Correlations of Cognitive Indices with WRAT and Teacher Achievement Ratings

	FSIQ	GCI	FSBC	BSBC	Mean	S.D.
WRAT						
Reading	.66	.66	.71	.67	117.0	21.0
Spelling	.63	.64	.71	.71	109.1	18.6
Arithmetic	.45[a]	.56	.60	.62	107.0	11.7
Total[b]	.65	.68	.74	.72		
Teacher Rating						
Reading	.76	.73	.77	.74	3.0	1.4
Spelling	.70	.53	.73	.69	2.8	1.1
Arithmetic	.62	.54	.69	.69	2.9	1.1
Total[b]	.73	.64	.77	.74		

Note: n = 30. FSIQ = Full Scale IQ (Wechsler Intelligence Scale for Children - Revised); GCI = General Cognitive Index (McCarthy Scales for Children's Abilities); FSBC = Full Scale Broad Cognitive Ability, BSBC = Brief Scale Broad Cognitive Ability (Woodcock-Johnson Tests of Cognitve Ability); WRAT = Wide Range Achievement Test. WRAT coefficients are Pearson correlations while Teacher Ratings are partial correlations adjusted for age.

[a] $p < .01$ for FSIQ with WRAT Arithmetic; $p < .001$ for all other correlation coefficients.

[b] Totals were computed by summing subscale scores.

Table 3

Partial Correlations between WRAT and Teacher Rating Scales

		Teacher Ratings			
		Reading	Spelling	Arithmetic	Total[b]
WRAT	Reading	.81[a]	.78	.71	.81
	Spelling	.77	.79	.72	.80
	Arithmetic	.54	.58	.63	.61
	Total[b]	.81	.81	.77	.84

Note: n = 30. WRAT = Wide Range Achievement Test.

[a] $p < .001$ for all correlation coefficients.

[b] Totals were computed by summing subscale scores.